

## 프로젝트 보고서

**Title : MEPFIL USP 1 값 예측(선형회귀)**

**과목 : 머신러닝 프로그래밍**

**지도교수 : 김성수 교수님**

**작성일 : 2023-06-05**

**학번 : 2023254002**

**성명 : 이민수**

1. 서론.....	3
1.1 주제 선정.....	3~4
1.2 주제 목적.....	4~5
1.3 선형회귀모델의 소개.....	5
1.4 다층 퍼셉트론의 소개.....	5~7
2. 분석 및 방법.....	8
2.1 Flow Chart.....	8
2.2 분석 및 방법.....	9~17
3. 결과.....	18
3.1 상관 관계 분석.....	18~19
3.2 선형 회귀 분석.....	19~21
3.3 다층 퍼셉트론 분석.....	22
4. 결론.....	23~24
5. 참고 문헌.....	25

# 1. 서론

## 1.1 주제 선정

봉합사는 의료 수술에 필요한 재료로써 고순도 고분자가 필요하다. 또한 안정성이 필요함으로서 염증이나 독성 반응이 없어야 하고 보관이나 소독이 용이해야 한다. 목적에 맞춰 다양한 형태로 가공할 수 있어야 하며 생체적합성, 생분해성 등 특성을 갖춰야 한다. 봉합사는 생체에 분해 흡수성에 따라 흡수성 봉합사와 비흡수성 봉합사로 나눌 수 있다. 최근에 봉합사는 단순히 조직 접합용 아니라 미용, 성형이나 한방에도 인기를 많이 끌고 있다. 특히 미용 성형 분야에 리프팅실로 많이 사용하고 있다. 리프팅실을 피하조직에 삽입하여 피부를 당기면서 주름 개선 효과를 나타낸다. 그렇기에 봉합사는 인장력과 탄성력 그리고 유연성 모두 좋아야 하며 적절한 분해기간도 필요하다. 이들은 PGA(Polyglycolide)로 제작한다. 아직까지 많은 연구 개발이 지속적으로 이어지고 있으며, 이는 생분해성 고분자 재료가 일정한 기간이 지나면 체내에 물과 이산화탄소로 분해되어 인체에 독성이 없다는 장점이 있고, 또한 생체적합성을 지니고 있어 인체에 거부반응을 생기지 않는다.

### 생분해성 봉합사



(주)메타바이오메드는 이처럼 생분해성 봉합사인 자사 브랜드(MEPFIL, MEPFIL-LAC, MEPFIL-QUICK, MEPFIL-LAC QUICK, MEPFIL-II 의 총 5 가지 제품)가 있으며, 이를 생산하는 Cleanroom 시설을 보유중에 있다. 생분해성 봉합사를 제조하는데 있어 화학적 반응과 기계적인 요소에 따라 공정이 나뉘며 이 또한 공정에 따른 수율(불량율)이 존재 한다. 생분해성 봉합사의 인장 강력(gden)의 수율에 크게 영향을 미치는 공정인 멀티 공정을 분석하고자 한다.

멀티 공정에서 생산 조건으로는 온도(Temp), 압력(Press), 속도(Speed) 로 설정되며, 그 중 인장 강력(gden)은 합격/불합격을 확인하는 수치로써 (주)메타바이오메드의 회사 수익(수율 (합격/불합격))과 밀접한 관련이 있는 중요한 요소이다. 이러한 수율(합격/불합격)을 향상하거나 향상시키거나 현재의 수율을 유지하기 위해서 (주)메타바이오메드에서는 다양한 공정 특성을 분석하고 모형화하고 있다. 그러나 생산의 경우 매우 많은 공정변수 및 장비 변수에 장비 변수에 영향을 받아, 수율(합격/불합격) 요인을 파악하기 매우 어렵다. 따라서, 단순한 통계적 분석이나 경험적 기술로는 합격/불합격에 따른 수율을 향상하는데 한계를 지니고 있으며, 다양한 변동 요인을 도출하여 공정 최적화를 이루어 내야 한다.

## 1.2 주제 목적

앞서 언급한 화학적 반응으로 생분해성 봉합사가 생산된다. 국내 생분해성 봉합사 생산업체로는 삼양사 및 (주)메타바이오메드 이 두 기업으로 전무하다. 그만큼 화학적 반응에 따른 제품의 물성 안정화가 어렵고 제조 환경, PGLA 의 상태, 각 공정에 영향을 끼치는 기계 및 작업자등 여러가지 변수에 따른 기술력 난이도에 큰 어려움이 따른다. 그렇기에 본 저자는 사람이 아닌 기계적 학습을 통하여 위 같은 문제를 해결하고자 한다.

본 연구의 목적은 다음과 같다. 첫째, 회귀분석을 활용해 온도(Temp), 압력(Press), 속도(Speed)에서 특성별 중요도를 선정하고 이를 통해 생분해성 봉합사 제품의 어떤 특성이 인장 강력(gden)의 주요한 영향을 미치고 있는지 파악한다. 둘째, 선형회귀분석(linear regression)을 이용하여 MSE(평균제곱오차) 값을 확인하고 Best Model 을 찾는다. 셋째, 결과를 통하여 특성별 중요도가 인장 강력(gden)에 끼치는 영향을 확인한다.

본 연구의 이후 구성은 다음과 같다. 2 장에서는 전체적인 분석 및 방법론에 대해

상세히 서술하였으며 3 장에서는 이를 적용한 결과에 대해 서술하였다.  
마지막으로 4 장에서는 결론을 서술하였다.

### 1.3 선형회귀모델의 소개

통계학에서 선형 회귀(linear regression)는 종속 변수  $y$ 와 한 개 이상의 독립 변수 (또는 설명 변수)  $X$ 와의 선형 상관 관계를 모델링하는 회귀분석 기법이다. 한 개의 설명 변수에 기반한 경우에는 단순 선형 회귀(simple linear regression), 둘 이상의 설명 변수에 기반한 경우에는 다중 선형 회귀라고 한다.

선형 회귀는 선형 예측 함수를 사용해 회귀식을 모델링하며, 알려지지 않은 파라미터는 데이터로부터 추정한다. 이렇게 만들어진 회귀식을 선형 모델이라고 한다.

선형 회귀는 깊이있게 연구되고 널리 사용된 첫 번째 회귀분석 기법이다.<sup>[3]</sup> 이는 알려지지 않은 파라미터에 대해 선형 관계를 갖는 모델을 세우는 것이, 비선형 관계를 갖는 모델을 세우는 것보다 용이하기 때문이다.

선형 회귀는 여러 사용 사례가 있지만, 대개 아래와 같은 두 가지 분류 중 하나로 요약할 수 있다.

- 값을 예측하는 것이 목적일 경우, 선형 회귀를 사용해 데이터에 적합한 예측 모델을 개발한다. 개발한 선형 회귀식을 사용해  $y$ 가 없는  $x$  값에 대해  $y$ 를 예측하기 위해 사용할 수 있다.
- 종속 변수  $y$ 와 이것과 연관된 독립 변수  $x_1, \dots, x_p$ 가 존재하는 경우에, 선형 회귀 분석을 사용해  $x_j$ 와  $y$ 의 관계를 정량화할 수 있다.  $x_j$ 는  $y$ 와 전혀 관계가 없을 수도 있고, 추가적인 정보를 제공하는 변수일 수도 있다.

### 1.4 다층 퍼셉트론의 소개

MLP 란 여러 개의 퍼셉트론 뉴런을 여러 층으로 쌓은 다층신경망 구조 입력층과 출력층 사이에 하나 이상의 은닉층을 가지고 있는 신경망이다. 인접한 두 층의 뉴런간에는 완전 연결 => fully connected 된다.

다층 뉴런은 왜 필요할까? 복잡한 패턴 분류를 위해서는 입출력 간의 복잡한 변환 구조가 필요하다. 사용하는 뉴런의 수를 늘리고 층을 추가하여 복잡한

구조의 의사결정 경계를 생성할 수 있다. 단일뉴런 (퍼셉트론)으로는 선형분리 가능한 경계선만 생성가능함. 두개의 뉴런을 결합함으로써 XOR 과 같은 비선형 분리가 가능한 결정선을 생성할 수 있다. 뉴런을 추가함으로써 다각형 모양의 경계선을 생성할 수 있다. => 복잡한 데이터의 분류가 가능해진다! 일반적으로 다층 신경망은 (MLP 는) 2-3 개 계층으로 구성된다. MLP 의 구조는 입력층, 은닉층, 출력층의 전방향 신경망이다. (은닉층은 하나 이상)

### 1. 활성화 함수

퍼셉트론에서는 step 함수 (계단 함수) 를 활성화 함수로 사용, MLP 에는 다양한 비선형 함수들을 활성화 함수로 사용 한다.

인공 신경망을 통과해 온 값을 최종적으로 어떤 값으로 만들지 아래와 같이 설명 드린다.

- 1) step function : 임계값을 미리 설정하고 (threshold value) 임계값 보다 크면 1, 아니면 0
- 2) Sigmoid function : [0,1] 사이의 값
- 3) Linear function :  $y = x$
- 4) tanh function : [-1, 1] 사이의 값
- 5) ReLu function : 양수이면 그대로 , 음수이면 0
- 6) Leaky ReLU
- 7) ELU

2. 손실 함수 : 전체 오차는 목표 출력값에서 실제 출력값을 빼서 제공한 값을 모든 출력 노드에 대하여 합한 값이다 => 평균제곱법(MSE)

3. 역전파 알고리즘 (EBP : Error BackPropagation) : 입력이 주어지면 순방향으로 계산하여 출력을 계산한 후에 실제 출력과 우리가 원하는 출력 간의 오차를 계산한다. 오차를 역방향으로 전파하면서 오차를 줄이는 방향으로 가중치를 변경한다.

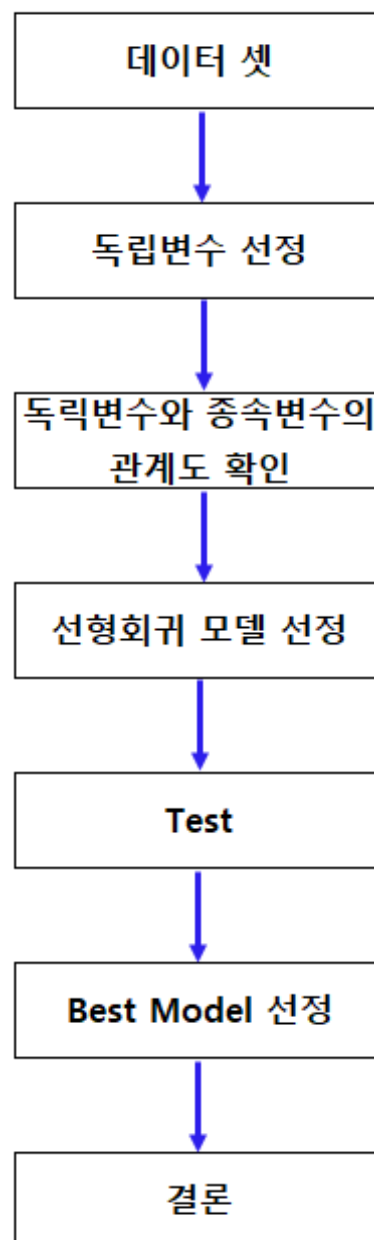
4. 다층 신경망의 활용: 다양한 지도 학습 문제에 적용 가능, MNIST 데이터셋에 적용, 스팸 여과기 만들기.

## 2. 분석 및 방법

### 2.1 Flow Chart

Pig1. Flow Chart 그대로 순서대로 데이터 셋, 변수 선정, 관계도 확인, 모델 선정, Test, Best Model 선정, 결론 순으로 진행을 실시하였다.

Pig1. Flow Chart



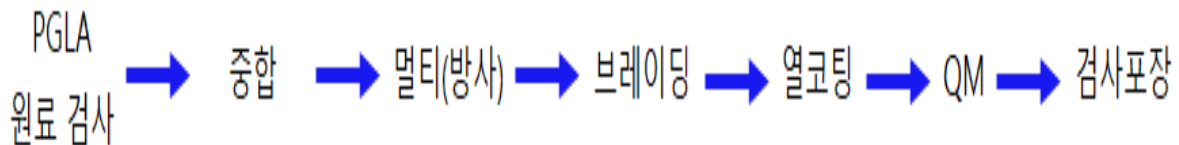


## 2.2 분석 및 방법

### 2.2.1 멀티(방사) 공정 선정 이유

생분해성 봉합사를 제조하는 공정은 아래 Pig2. 제조 흐름도와 같다.

Pig2. 제조 흐름도



#### 1. PGLA 원료 검사

PLGA 원료를 자가 중합 또는 외부 업체로부터 원료를 수입받는다. 이 때 PGLA의 성적서(입도분석기)를 확인하거나 자가 입도분석기를 활용하여 자사 내 기준치에 만족하는지 확인한다.

#### 2. 중합

원료 검사에서 합격된 원료를 가지고 중합(특수 공정)을 실시한다. 해당 공정에 관하여 자세하게 말씀드리기가 어려우나 파우더 형태의(PGLA)를 결정체를 가진 특수한 모양으로 변형시키는 공정이다.

#### 3. 멀티(방사)

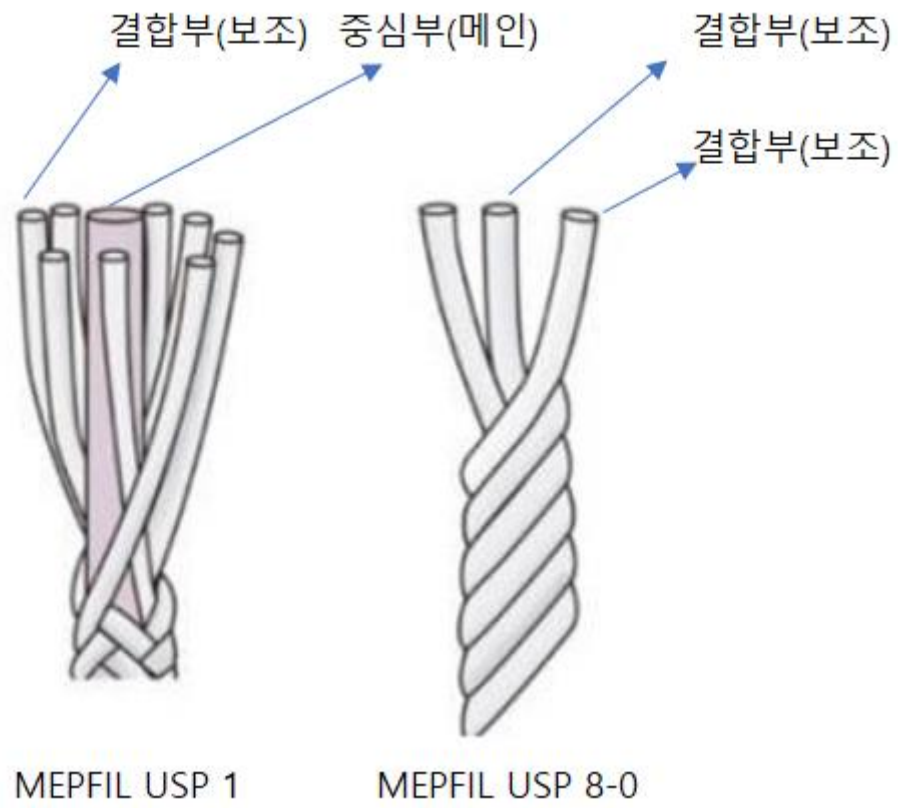
중합에서 제조된 결정체를 방사기에 투입하여 제품별, SIZE 별 방사하는 공정이다. 이때, 결정체를 녹여 방사를 위한 온도(Temp), SIZE를 위한 압력(Press), 방사되는 봉합사의 뽑는 속도(Speed)가 주요 요인이다.

봉합사의 기본이 되는 매우 얇은 실들을 결합하여 최종 봉합사를 만들게 되는데, 이때 인장 강력(gden)에 가장 중요한 결합부(보조), 중심부(메인)을 제조하는 공정이다. 이 결합부(보조), 중심부(메인)의 상태에 따라 인장 강력(gden)의 값에 영향이 생긴다. 해당 설명 사진은 Pig3. 봉합사 결합 상태로 표시하였다.

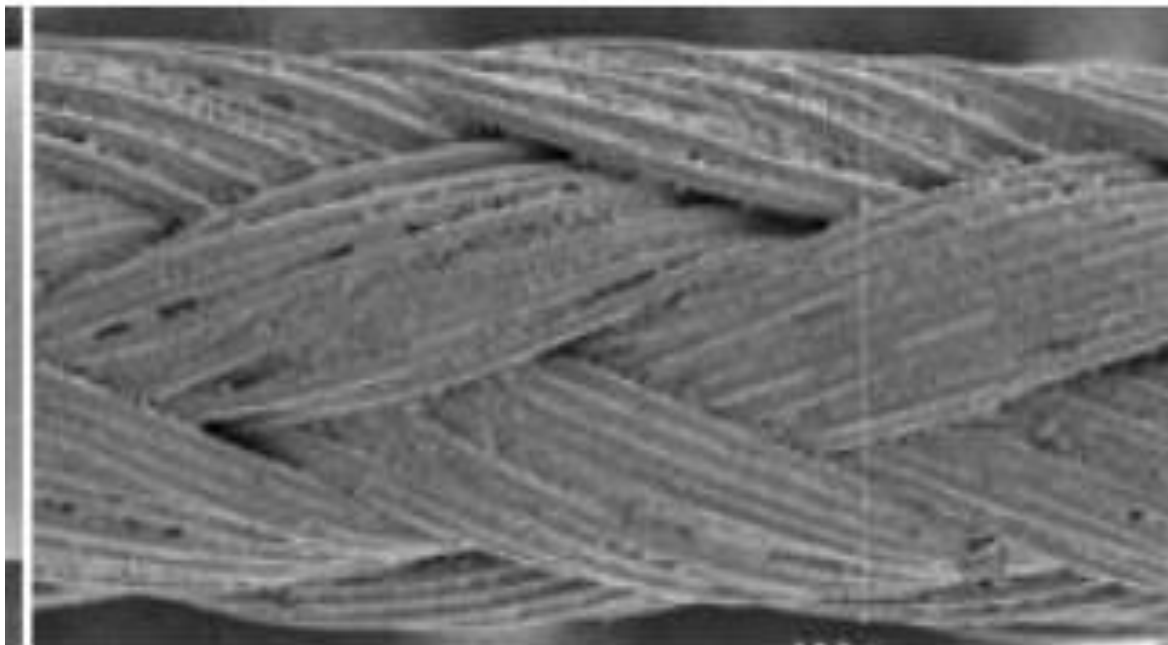
Pig4. 봉합사 현미경 사진같이 현미경으로 확대하여 봤을 때, 무수한 멀티(방사)실들이 결합되어 있는 모습을 볼 수 있다.

멀티(방사)에서 제조한 실들을 결합하여 생분해성 봉합사를 제조하는데, 그에 관한 설명은 다음 공정인 브레이더 공정에서 자세히 설명하도록 한다.

Pig3. 봉합사 결합 상태



Pig4. 봉합사 현미경



#### 4. 브레이딩

멀티(방사)에서 나온 봉합사를 SIZE 규격에 맞게 8 축, 16 축등으로 브레이딩(꼬임)을 실시하여 제품을 제조 한다.

#### 5. 열코팅

브레이딩(꼬임)이 끝난 반제품을 열을 가하는 공정이다. 알다 싶이 현재 브레이딩 공정에서 여러 가닥의 실들을 결합하여 하나의 실로 만들어져 있는 상태이다.

문제는 이 실들에 곳곳에 공기층이 들어가 있어 하나로 응집되어 있지 않기 때문에 인장 강력(gden)이 낮을 것이다.

이것을 해결하기 위하여 열(해당 제품,SIZE 별 기준)을 가해 응집될 수 있도록 열 공정을 거친다.

그 이후, 코팅 공정에서 봉합사가 쉽게 묶이거나 풀릴 수 있도록 외벽에 코팅을 칠하는 공정으로 마무리 하게 된다.

#### 6. QM

앞선 공정들이 모두 정상적으로 이루졌다면, QM 에서 생분해성 봉합사(KGMP, EU, FDA 의료기기법)에 기준치에 만족하는지 검사를 실시한다. 가장 중요한점은 생분해성 봉합사의 인장 강력(gden)의 값이다.

7. 모두 합격된 제품을 검사포장을 걸쳐 완제품으로 끝나게 되며, 그 이후 공정은 출하입고 및 출하대기로 옮겨져 추후 고객에게 납품이 된다.

먼저 멀티(방사) 공정을 선택한 사유는 아래와 같다.

- 모든 공정 중 인장 강력(gden)을 측정하는 공정은 멀티(방사) 공정 뿐이다 (멀티(방사) 공정 자가 검사 실시)
- 멀티(방사) 공정에 제조된 결합부(보조), 중심(메인) 모두 하나 하나가 합격 되어야만 정상 제품으로 제조가 된다.
- 그렇기에 멀티(방사)에서 제조되는 실들의 인장 강력(gden)의 값이 제조 되는 변수(온도, 압력, 속도)에 어떠한 영향이 있는지 가장 먼저 파악하는 것이 중요하다.

### 2.2.2 제품 선정

제품 선정은 MEPFIL, MEPFIL-LAC, MEPFIL-QUICK, MEPFIL-LAC QUICK, MEPFIL-II 중 2022 년 기준 가장 많이 생산된 것과 가장 많이 매출을 올린 제품으로 선정하였다. 2022 년 Data 확인 결과 제품은 MEPFIL 이며 SIZE 는 USP 1 으로 결정하였다. MEPFIL USP 1 은 (주)메타바이오메드의 주력 상품 및 인장 강력(gden)의 편차가 그리 많지 않은 점과 고객불만 접수 건이 다른 제품들에 비해 월등히 낮다는 이유도 있다. 고객불만 접수가 낮다는 것은 제품 물성이 안정화가 높다는 뜻이고, 이는 특성치(독립변수, 종속변수)를 찾을 수 있을 것이라 판단되기 때문이다. 물성의 안정화의 편차가 높다면 특성치(독립변수, 종속변수)를 찾기 어려울 것이다.

### 2.2.3 Data 수집

선정된 MEPFIL USP 1 을 제조되는 멀티 방사의 2022 년 1 월 1 일 ~ 2022 년 12 월 31 일까지의 제조 Data 를 확보하고 독립변수 및 종속변수를 선정하여 csv file 로 재 정리 진행하였다. Data 정리결과 총 3,719EA 로 정리할 수 있었다. 분석결과 만족하지 못한다면 3 년치(2022 년, 2021 년, 2020 년도 멀티 MEPFIL USP 1 Data)를 수집하여 사용하도록 한다.

### 2.2.4 변수 정의 및 선택

생산 조건은 Pig5. 멀티 MEPFIL USP 1 조건과 같다. 이 조건을 Table1. 조건 설명으로 상세히 기재 및 특성치(변수)를 선정하였다.

POP 도입으로 TSSR = MEPFIL 임을 강조하며, 01 은 USP 1 임을 말씀드린다.

Pig 5. 멀티 MEPFIL USP 1 조건

그룹 LOT (POP)		TSSR12201G001				방사기	작업일자	시험의뢰	FDY강도 Target		달성율	Denier	1km/g	
(구) LOT						E	2022-01-26 (수)	S2-96	목표	8.4	96%	91.0	양품	229.4
		수율	차이	0.3	4%			core	246.0					
제품명(선택)		MEPFIL USP 1						92.3%	투입량	43,492 m		불량수량	3,365 m	
지관 No	섬도 (Denier)	편차 (Denier)	강도 (g/den)	신도 (%)	강력 (gf)	무게 (g)	Chip 이력 사항		권취 시간	Melt temp (온도℃)	Pack Press (psi)	1 G/R Speed (mpm)	권자 상태 (내,외층)	
							LOT / MI / 수분	사용량(g)						
Avg.	90.7	0.4	8.1	31.3	731		총 9,977 g			261.0	1028	420		
R.	0.8		0.6	4.4	64		↑ TotalL 사용량(g)			0.8	30	0		
1	90.5		8.136	30.36	737	1,143	L1-205	9,977	0:40	261.1	1,010	420	양호	
2	90.5	0.0	8.130	33.27	735	1,154	22.2/20		1:40	261.1	1,020	420	양호	
3	91.0	0.5	8.366	31.52	762	1,151	2021-10-21		2:40	261.1	1,027	420	양호	
4	91.0	0.0	7.729	30.49	704	1,151			3:40	261.1	1,025	420	양호	
5	91.0	0.0	7.982	31.99	727	1,154			4:40	261.1	1,040	420	양호	
6	90.6	0.4	8.317	31.71	753	1,149			5:40	261.2	1,036	420	양호	
7	90.2	0.4	7.743	31.87	698	1,150			6:40	261.1	1,033	420	양호	
8	90.6	0.4	8.060	28.87	730	1,153			7:45	260.4	1,033	420	양호	

Table1. 멀티 MEPFIL USP1 조건 설명

섬도(Denier)	-
편차(Denier)	PGLA Chip
<b>강도(gden)</b>	<b>PGLA Chip 을 녹여 나오는 제품의 인장 강력 값</b>
신도(%)	-
무게(g)	PGLA Chip 을 녹여 나온 제품의 무게
Chip 이력사항	PGLA Chip 이력사항
권취시간	-
<b>Melt Temp</b>	<b>PGLA Chip 을 녹이는 기계 온도</b>
<b>Pack Press</b>	<b>PGLA Chip 을 녹여 생산하는 노즐 압력</b>
<b>1 G/R Speed</b>	<b>PGLA Chip 을 녹여 생산하는 속도</b>

MEPFIL USP 1 을 생산하는데 필요한 작업표준서(쥬메타바이오메드 문서)을 1 차로 확인하였으며, 이를 근거로 2 차로 위 표를 보아 MEPFIL USP 1 을 생산하는데 있어 필요한 온도(Temp), 압력(Press), 속도(Speed) 임을 확인하였다. 독립변수로 온도(Temp), 압력(Press), 속도(Speed) 이 세 가지를 선정하였고 종속변수로는 인장력(gden)을 선정하였다.

마지막으로 위에 언급한 제품 선정, Data 확보, 변수 정의에 따른 파이썬 종류 및 알고리즘을 정리한 내용을 Table2. Data Sampling 으로 표시하였다.

Table2. Data Sampling

제품명	MEPFIL USP1
Data 기간	2022 년 1 월 ~ 12 월 생산 Lost
Data 수	3,719EA
파이썬	구글 코랩
선형회귀	Linear, Ridge, Lasso, Elastic Net, RANSAC
X(독립변수)	온도(Temp), 압력(Press), 속도(Speed)
Y(종속변수)	인장 강력(Melt gden)

1	Melt gden	Temp	Press	Speed	3701	7.896	269.9	1597	470
2					3702	8.037	269.8	1603	470
3	8.136	261.1	1010	420	3703	8.373	270.1	1605	470
4	8.130	261.1	1020	420	3704	7.996	270.0	1612	465
5	8.366	261.1	1027	420	3705	8.095	270.0	1610	465
6	7.729	261.1	1025	420	3706	7.901	270.0	1634	465
7	7.982	261.1	1040	420	3707	8.038	270.0	1631	465
8	8.317	261.2	1036	420	3708	7.910	270.0	1590	475
9	7.743	261.1	1033	420	3709	7.688	270.0	1572	475
10	8.060	260.4	1033	420	3710	7.774	269.9	1574	475
11	7.726	260.0	1087	495	3711	7.793	270.0	1575	475
12	7.766	259.9	1094	495	3712	7.726	270.0	1583	475
13	7.879	259.9	1107	495	3713	8.017	270.0	1588	475
14	7.582	259.9	1108	495	3714	8.358	270.0	1600	475
15	7.635	260.0	1109	495	3715	7.688	270.0	1613	475
16	8.287	260.4	1108	490	3716	7.887	270.0	1603	470
					3717	7.926	270.0	1604	470
					3718	8.085	270.0	1608	470
					3719	7.580	270.1	1611	470

## 2.2.5 변수 배제

- 생분해성 봉합사는 Cleanroom 에서 제조된다. Cleanroom 이란

청정실(淸淨室) 또는 클린룸(cleanroom)이란 공기 중 입자들의 집중이 제어되도록 제어되었으며, 먼지의 발생, 유입, 유보가 최소화되도록 되어 있고, 온도, 습도, 그리고 압력이 요구에 따라 제어되는 방을 의미한다.

이 글에서 말하는 입자(Particle)는 먼지 뿐만 아니라 세균, 바이러스, 금속 가루, 사람 몸에서 나오는 피부 세포 등등 공기 중에 떠다닐 만한 모든 입자를 이르는 것이다. 그리고 이 글에서는 그중에서  $0.1\mu\text{m} \sim 10\mu\text{m}$  수준의 입자를 일컫는다.

특히 이 수준으로 쪼개진 입자들은 공기 중에 부유하는 것이 가능하고, 특히  $0.5\mu\text{m}$  이하의 입자들은 중력의 힘으로는 제대로 침강하지 않으므로, 이러한 입자들이 문제가 된다면 반드시 클린룸을 설치하여야 한다.

2 차 세계 대전 이후 많은 공산품이 쏟아지고 생산성에 관심이 높아지는 가운데, 많은 불량 혹은 고장의 원인으로 입자(Particle)가 지목되었다. 특히 세균의 발견과, 의료기기의 생산에서 클린룸의 중요성은 매우 높아지게 되었다.  $1\text{m}^3$  당 일반적인 입자는 35,000,000 개로, 이러한 입자들이 부유하는 곳에서의 생산은 피하는 것이 필수가 되었다. 특히 의료기기의 경우 공정이 더욱 미세해짐에 따라 이러한 조치는 더욱 필요해졌다. 특히 노광(Photo) 공정에서 이러한 입자는 노광을 방해하는 주요 입자로 지목되었고, 극도의 청정도를 요구하게 된다.

(주)메타바이오메드의 의료기기 Cleanroom 관리 기준 항목은 아래와 같다.

온도 :  $1^{\circ}\text{C} \sim 30^{\circ}\text{C}$

습도 : 30% RH ~ 70% RH

Particle(부유입자) : 3.530.000 이하 (ISO Class 8)

차압 :  $+0.4\text{mmH}_2\text{O} \sim 5.0\text{mmH}_2\text{O}$

(주)메타바이오메드는 일상점검으로 온도, 습도, 차압을 매일 하루 2 회 측정하고 있으며, 이는 모두 기준치에 만족함을 확인하였다. 또한, Particle(부유입자)는 매 1 회 Cleanroom Validation 을 통하여 기준치에 만족함을 확인하였고, 매년 측정치를 확인하여 변수 발생시 즉각 해결하고 있다. 그렇기에 Cleanroom 관리 기준은 이번 실험에서 변수를 제외하도록 진행하였다.

- 별도로 인장 강력(gden)을 측정한 INSTORN 장비는 특성치에 고려하지 않았다. 멀티(방사) 및 QM 에서 사용중인 INSTORN 장비는 매번 교정을 실시하고 있으며,



교정값 확인 결과 인장 강력(gden)에 큰 영향을 끼치지 않을 것이라 판단하였기 때문이다.

또한 실제 사용한 Data 는 QM Data 가 아닌 멀티(방사) 공정에서 자가 측정된 Data 이기 때문에 변수에서 제외시켰음을 알려 드린다.

- 각 방사기(멀티 MEPFIL USP 1 을 제조하는 기계)는 A, B, C 가 있지만 이 역시도 설치 적격성 평가(Installation Qualification, IQ), 운전 적격성 평가(Operational Qualification, OQ), 성능 적격성 평가(Performance Qualification, PQ) 문서 결과 차이가 미미 하기에 동일시하여 진행하기로 결론 내렸다.

## 3. 결과

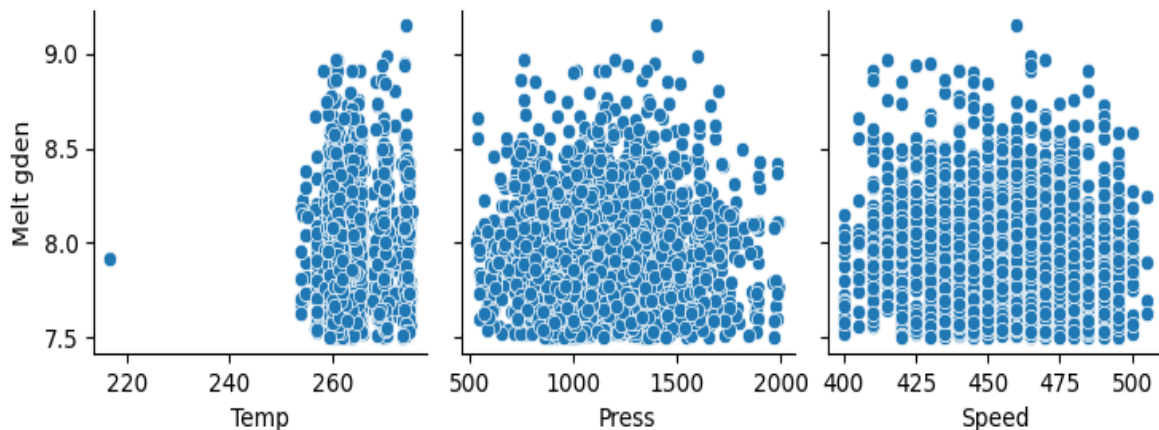
### 3.1 상관 관계 분석

온도(Temp), 압력(Prees), 속도(Speed) 및 인장 강력(gden) 값을 포함하여 변수에 대한 기술 통계 요약을 진행 하였다. 자세한 내용은 아래와 같으며 Fig6. 상관 관계 분석을 참고하면 된다.

- 회귀 분석으로 구한 절편 값은 7.969475114571974 이다.
- 온도(Temp)에 대한 계수는 0.0022050614808892402 이며, 이는 온도(Temp) 단위 증가시 인장 강력(gden) 값이 약 0.0022 증가할 것으로 예상되는 수치이다.
- 압력(Prees)에 대한 계수는 1.708134945007534e-05 이며, 이는 압력(Prees) 단위 증가시 인장 강력(gden) 값이 약 1.708e-05 증가할 것으로 예상되는 수치이다.
- 속도(Speed)에 대한 계수는 -0.0014201809196507484 이며, 이는 속도(Speed) 단위 증가시 인장 강력(gden) 값이 약 1.708e-05 증가할 것으로 예상되는 수치이다.
- 종속 변수인 인장 강력(gden)을 기준으로 온도(Temp)는 +0.002, 압력(Prees)은 -1.708e-05, 속도(Speed)는 -0.001 의 값을 확인하였다. 결론적으로 온도(Temp), 압력(Prees), 속도(Speed) 이 세 가지 독립 변수 중 온도(Temp)가 압력(Prees), 속도(Speed)에 비해 종속 변수인 인장 강력(gden)과 상관관계가 그나마 높다는 것을 확인하였고, 이를 근거로 선형회귀가 아닌 비선형회귀 알고리즘을 사용하기로 결정하였으며, 독립 변수로 더 이상 압력과 속도는 설정하지 않고 온도(Temp)와의 상관관계를 찾는 테스트를 진행하였다.

Fig6. 상관 관계 분석

Intercept: 7.969475114571974  
Coefficient for Temp: 0.0022050614808892402  
Coefficient for Press: 1.708134945007534e-05  
Coefficient for Speed: -0.0014201809196507484  
Correlation between Temp and Melt gden: 0.025167487412849406  
Correlation between Press and Melt gden: -0.014743259418231621  
Correlation between Speed and Melt gden: -0.10519655689041978



### 3.2 선형 회귀 분석

독립 변수로 온도(Temp)와 종속 변수인 인장 강력(gden)을 활용한 선형회귀 모델 및 Best Model 을 찾는다. 선형회귀 모델은 Linear, Ridge, Lasso, Elastic Net, RANSAC 모델로 총 5 가지를 선정하였으며, 이 중 MSE 의 값으로 Best Model 로 선정하기로 하였다. 결과는 아래와 같다. (Pig7. MSE Data & Pig8. Temp-gden 그래프 참조)

- Linear Regression : 0.06999095600971
- Ridge Regression : 0.06999095983446414
- Lasso Regression : 0.07016727998272637
- Elastic Net Regression : 0.07007938669424527
- RANSAC Regression : 0.07242690979625485

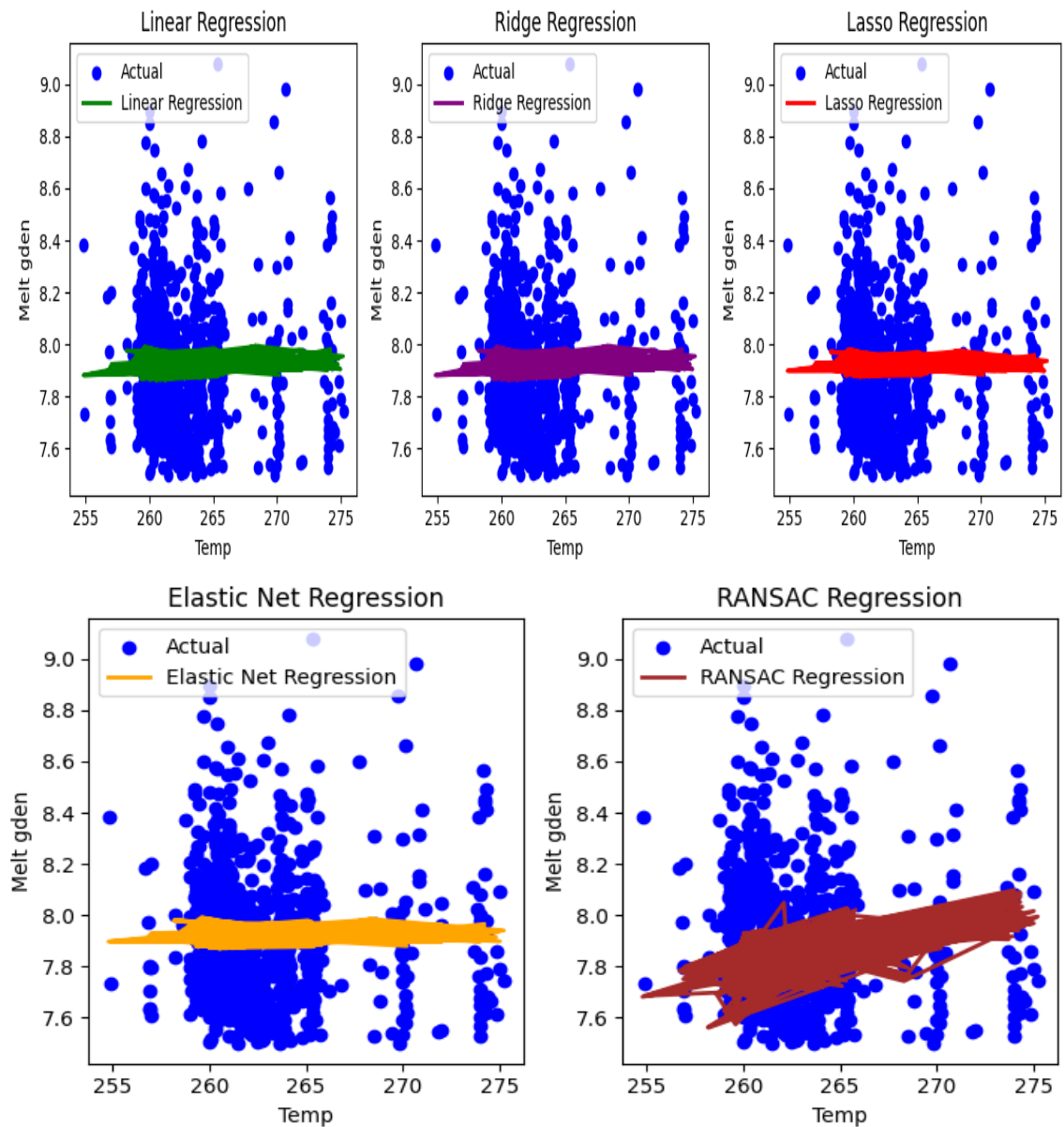
인장 강력(gden)의 값을 예측하기 위한 최적의 모델은 MSE 값이 가장 낮은 Linear Regression(0.06999095600971)로 파악된다.

Pig7. MSE Data

Mean Squared Error:  
Linear Regression: 0.0699909595600971  
Ridge Regression: 0.06999095983446414  
Lasso Regression: 0.07016727998272637  
Elastic Net Regression: 0.07007938669424527  
RANSAC Regression: 0.07242690979625485

Best Model :  
Model: Linear Regression  
MSE: 0.0699909595600971

Pig8. Temp-gden 그래프



Best Model 로 선정된 Linear 로 R2(결정계수) 및 CVMSE(Cross Validated MSE)의 값을 재차 확인하도록 코드를 재실행 하였고 결과는 Fig9. R2&CVMSE 와 같으며 값 밑에 간략한 해석을 알려드리고자 한다.

- Mean Squared Error : 0.0699909595600971

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

굉장히 많이 사용하는 척도 중 하나인데, 제곱으로 인해 그 원래의 차이보다 다소 민감한 성능 평가가 될 수 있는 점을 고려하여 사용해야 한다. 평균제곱오차는 위 식과 같이 잔차제곱합(RSS)을 해당 데이터의 개수로 나누어서 구할 수 있다.

예측값 - 관측값(데이터값)의 제곱된 값의 평균을 구하는 것이다. 여기서 잔차의 제곱을 하는 이유는 잔차의 값이 음수가 될 수 있는 것을 방지할 수 있고, 제곱을 함으로써 오차의 민감도를 높이기 위함이다.

- R-squared : 0.019752105062800562

- X 와 Y 간의 상관관계가 클수록 R2 의 값은 1 에 가까워진다. - 즉, R2 값이 0 에 가까워 질수록 회귀선은 쓸모가 없고 값이 클수록( $R^2 \geq 0.65$ ) 쓸모있는 회귀식이라고 본다.

- 총변동중에서 회귀선에 의해 설명이 되는 변동이 차지하는 비율을 의미한다.

- Cross-Validated MSE : 0.07107441387897284

Fig9. R2&VMSE

Mean Squared Error (MSE): 0.0699909595600971

R-squared (R2): 0.019752105062800562

Cross-Validated MSE: 0.07107441387897284

MSE 결과 값은 0.06 으로 좋은 모델임을 증명하였지만, R2 의 결정계수는 0.01 로 현저히 낮은 값이 나왔다.

### 3.3 다층 퍼셉트론 분석

추가로 다층 퍼셉트론(MLP)를 이용하여 분석을 실시하였다. 결과에 대한 요약은 아래와 같으며 결과 값은 Pig10. MLP 결과 로 표시하였다.

- MLP Regressor (ReLU): 0.5186476877249401

- MLP Regressor (Sigmoid): 0.07130011010383804

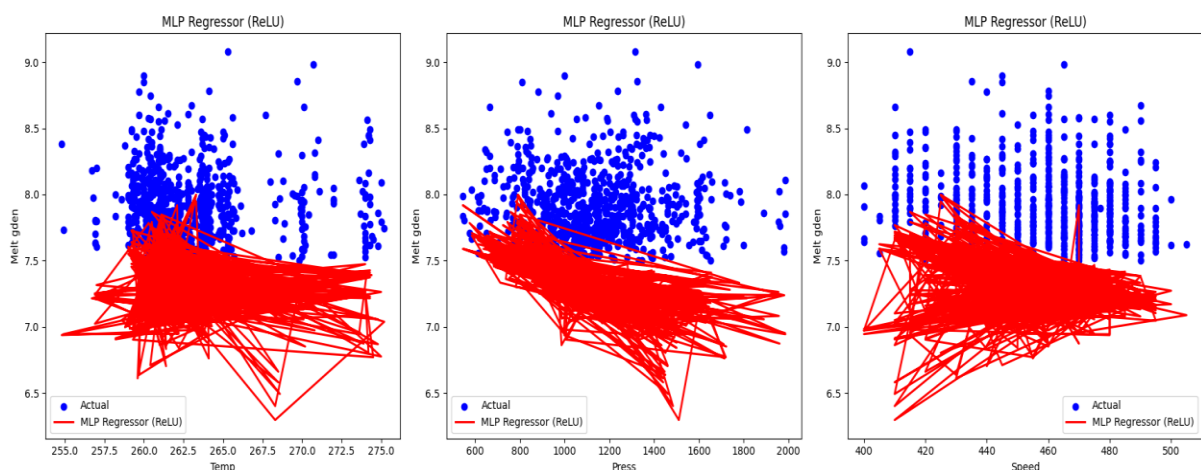
Relu 를 사용한 MLP 의 결과 MSE 는 0.5186476877249401 입니다. Sigmoid 의 MLP 의 결과 MSE 는 0.07130011010383804 입니다. Relu 모델은 0.518 의 MSE 를 달성한 반면 Sigmoid 모델은 0.071 MSE 를 달성했습니다.

Sigmoid 모델에 의해 얻은 더 낮은 MSE 는 이 특정 데이터 세트에서 입력 기능과 대상 변수 간의 관계를 근사화 하는데 더 잘 수행되었음을 나타냅니다. 활성화 함수의 성능은 데이터 세트와 해결하려는 특정 문제에 따라 달라질 수 있다는 것을 다시 한번 생각해야 하며, 요약하자면 Relu 함수 보다 Sigmoid 함수가 이 문제를 대한 데이터의 기본 패턴을 분석하는데 더 나은 성능을 나타냄을 알 수 있다.

Pig10. MLP 결과

MLP Regressor (ReLU): 0.5186476877249401

MLP Regressor (Sigmoid): 0.07130011010383804



## 4. 결론

생산, 품질팀 담당자가 모여 나온 결과 값을 정밀히 검토 및 의견을 수립한 결과 (주)메타바이오메드 MEPFIL USP 1 의 적합/부적합을 나타내는 gden 의 값과 생산 조건으로 온도(Temp), 압력(Press), 속도(Speed)의 관계를 아래와 같이 결론을 내리게 되었다.

- 앞서 R2(상관계수)의 값이 0.01 로써 독립 변수와 종속 변수간 서로의 영향은 매우 미비한 것으로 결론 내렸다. 하지만, 관리 차원에서 작업 표준화에 맞춰 MEPFIL USP 1 생산이 설정된 값을 준수하여 기계를 운용해야 한다.
- 독립 변수 온도(Temp)는 종속 변수 인장 강력(gden) 값에 다른 독립 변수 압력(Press), 속도(Speed)에 비하여 상관 관계가 있지만, 이를 가지고 인장 강력의 값을 추정하기 매우 어려움을 확인 하였다.
- 멀티 공정 전 원료 입도 검사(입도분석기로 입도 분석) 및 중합 공정에서의 영향, 그리고 멀티 공정의 Cleanroom 상태 등 다양하고 넓게 모든 공정의 변수 및 특정치를 검토해볼 필요가 있다고 판단된다.  
입도 분석의 평준화를 위하여 기준을 재정립할 필요가 있으며, 이를 적용하기 위해 개발팀이 주관하여 제품 Test 를 먼저 선 진행하도록 한다. 그리고 양산화를 진행하여 추후 분석을 이어 나가도록 진행할 예정이다.
- 합성고분자인 PGA, PLLA 및 공중합체인 PLGA 을 이용하여 제조하는 생분해성 봉합사는 복잡한 제조 방법 및 그 제조 환경에 따라 인장 강력(gden)의 값을 결정한다.  
화학적 반응과 모든 제조 환경의 변수를 더 확장할 필요가 있다. 이를 위해서 광범위한 Data 분석가가 필요로 하며 이를 위해 필요시 TF 팀을 구성할 필요가 있다.

- 이번 과제를 통하여 비록 상관 관계를 찾지 못하였지만, 졸업 전까지 심도 깊게 다시 한번 자료 조사 및 Data 분석을 통하여 상관 관계를 찾도록 업무 진행 예정이다.



## 5. 참고 문헌

5.1 등록특허 10-1143420 생분해성 수술용 봉합사

5.2 머신 러닝 교과서 with 파이썬, 사이킷런, 텐서플로 개정 3 판

5.3 위키 백과사전